BOSTON UNIVERSITY

GRADUATE SCHOOL

Thesis

THE THEORY OF RANGE AND ITS APPLICATION TO QUALITY CONTROL

by

Sister Catherine Josephine Gillis, S.N.D.

(A. B., Emmanuel College, 1942)

submitted in partial fulfilment of the requirements for the degree of Master of Arts.

1946

1946

g i

Approved

by

First Reader.....Elmer B. Mode.....

Professor of Mathematics

Second Reader.....Ralph M. Johnson.....

Professor of Mathematics

OUTLINE

PAGE

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

NATURE OF RANGE AND SCOPE OF THE THESIS

Statistics, though not a new science, is a developing one. Quite naturally, then, many of its measures are still under investigation. One such measure, the theory of which has been recently developed and is, in fact, still in process of development, is the statistical measure of dispersion, the range.

I. NATURE OF RANGE

Definition. The range has been defined as ". . . the difference between the greatest and smallest members of a sample";[1] ". . . the difference between the highest recorded score and the lowest recorded score";[2] ". . . the difference between the maximum and minimum observations".[3] Representing the distance between extreme observations, it is the simplest possible measure of a group of measures. It

---

[1]L. H. C. Tippett, The Methods of Statistics (second edition, revised; London: Williams and Norgate, Ltd., 1937), p.31.

[2]C. H. Richardson, An Introduction to Statistical Analysis (enlarged edition; New York: Harcourt, Brace and Company, 1941), p. 80.

[3]E. S. Pearson, The Application of Statistical Methods to Industrial Standardisation and Quality Control (London: British Standards Institution, 1935), p. 109.

gives a comprehensive value of the data in that it includes the limits within which all of the items occur.[4] Hence it appears to be the most natural index of dispersion.

Advantages and disadvantages. Nevertheless, it has been little used for purposes of comparison. The extreme ease with which it may be calculated and its very obvious interpretation which have led to its use in many industrial problems, are frequently more than offset by certain serious objections.[5] Determined by only the two extreme measures, it tells nothing of the form of the distribution within the range. A symmetrical and a J-type frequency curve might have the same value for the range. It tells nothing about the concentration of the measures about the center.[6] If either one (or both) of the extremes is an unusual occurrence it may have quite a disproportionate effect on the range.

Relation to σ. Moreover, the range varies for samples of different sizes taken from the same population, being smaller for smaller samples. For a normal population,

---

[4]Frederick E. Croxton and Dudley J. Cowden, Applied General Statistics (New York: Prentice-Hall, Inc., 1941), p. 236.

[5]G. Undy Yule and M. G. Kendall, An Introduction to the Theory of Statistics (eleventh edition, revised; London: Charles Griffin and Company, Ltd., 1937). p. 134.

[6]Richardson, loc. cit.

however, the mean range over k samples (k large) bears a fixed relationship to the more common measure of variation, the standard deviation of the population.[7] Nor does this ratio seem to be very sensitive to moderate changes in the form of distribution, so that it can be said that, in general, when the population form and sample size are constant, the mean range is proportional to the standard deviation.[8] It is from this property of the range, as well as from the fact that it is easily computed and easily understood, that its utility arises, its chief use up to the present being in its application to quality control.

II. SCOPE OF THE THESIS

__Purpose of the thesis__. Because of the growing importance of this phase of statistical work, a discussion of range theory seemed justifiable. How this theory may be utilized to simplify the numerical computations in control chart analysis has also been demonstrated.

__Division of the material__. This study, then, has been divided into two parts. Part I treats of the theory of range that has been developed up to the present. An attempt

---

[7] H. A. Freeman, __Industrial Statistics__ (New York: John Wiley & Sons, Inc., 1942), p. 131.

[8] Tippett, __loc. cit.__

has been made to collect all available material dealing with the distribution of range in general, and in particular with the distribution of range when the samples are drawn from a normal population. The method of moments has been studied as being the most useful. The calculation of percentage limits has been investigated with a view to using these findings in control chart analysis. In Part II the calculation of the standard deviation from the range has been considered and this estimate compared with those obtained by more rigorous methods. The control chart for sample ranges has been constructed together with that for sample deviations. The similarity between these two charts has been the basis for the conclusions of Part II.

PART I

THEORY OF RANGE

If a statistical measure is to be a useful tool, its nature must be thoroughly understood. The character of its distribution must be known. The distribution of range may be approached from two different standpoints. It may be developed from the distribution of K-order statistics or the moments of range may be computed with a view to finding the best-fitting curve. Both methods have been here considered.

I. SOME SPECIAL DISTRIBUTIONS

Distribution of any K-order statistics. Before considering the distribution of the range, it is necessary to determine the simultaneous distribution of any K-order statistics. The development given by Wilks[9] has been followed in this part of the discussion.

With $x_1, x_2, \ldots, x_n$, a sample of size n from a population with probability element $f(x)\,dx$, and with $x_1, x_2, \ldots, x_n$, arranged in ascending order of magnitude, he let $r_1, r_2, \ldots, r_k$, be k integers such that

[9] S.S.Wilks, Mathematical Statistics (Princeton: Princeton University Press, 1944), pp. 89-90.

$1 \leq r_1 < r_2 < \ldots < r_k \leq n$. His problem was to find the probability element of $x_{r_1}, x_{r_2}, \ldots, x_{r_k}$, or

$$f(x_{r_1}, x_{r_2}, \ldots, x_{r_k})\, dx_{r_1}\, dx_{r_2} \ldots dx_{r_k}.$$

He considered the sample to be covered by $2k+1$ intervals in such a way that each of the k elements x_{r_1}, $x_{r_2}, \ldots, x_{r_k}$ has its own interval, the other $k+1$ intervals covering the remaining $n - k$ elements of the sample, no two intervals overlapping. These intervals,

$$I_1, I_2, I_3, \ldots, I_{2k+1}, \text{ were}$$

$$(-\infty, x_{r_1}), (x_{r_1}, x_{r_1} + dx_{r_1}), (x_{r_1} + dx_{r_1}, x_{r_2}), \ldots,$$

$$(x_{r_k} + dx_{r_k}, \infty):$$

with

$$\int_{I_i} f(x)\, dx = q_i : \qquad (i = 1, 2, \ldots, 2k+1)$$

$$\left[\sum_{i=1}^{2k+1} q_i = 1\right]$$

The problem resolved itself into finding the probability (to terms of order $dx_{r_1}, dx_{r_2}, \ldots dx_{r_k}$) that if a sample of n elements is drawn from a multinomial population with classes $I_1, I_2, \ldots, I_{2k+1}$, then $r_1 - 1$ will fall in I_1, 1 element in I_2, $r_2 - r_1 - 1$ elements in I_3, 1 element in I_4, ..., $n - r_k$ in I_{2k+1}. From the multinomial law it follows that the probability of such a partition is

$$\frac{n!}{(r_1-1)\;1!\;\;(r_2-r_1-1)!\;1!\;\ldots(n-r_k)!}$$

$$q_1^{\;r_1-1}\quad q_2^{\;1}\quad q_3^{\;r_2-r_1-1}\quad q_4^{\;1}\;\ldots\;q_{2k+1}^{\;n-r_k}\qquad ,$$

which becomes, when the values of q_i are substituted,

$$\frac{n!}{(r_1-1)\;!\;(r_2-r_1-1)\;!\;\ldots\;(n-r_k)\;!}$$

$$\left[\int_{-\infty}^{x_{r_1}} f(x)dx\right]^{r_1-1}\left[\int_{x_{r_1}}^{x_{r_1}+dx_{r_1}} f(x)\;dx\right]\left[\int_{x_{r_1}+dx_{r_1}}^{x_{r_2}} f(x)dx\right]^{r_2-r_1-1}$$

$$\left[\int_{x_{r_2}}^{x_{r_2}+dx_{r_2}} f(x)dx\right]\;\ldots\;\left[\int_{x_{r_k}+dx_{r_k}}^{\infty} f(x)\;dx\right]^{n-r_k}.$$

Now, to within terms of order dx_{r_i},

$$\int_{x_{r_i}}^{x_{r_i}+dx_{r_i}} f(x)dx = f(x_{r_i})\;dx_{r_i}\quad\text{and}\quad\int_{x_{r_i}+dx_{r_i}}^{x_{r_i+1}} f(x)dx$$

$$= \int_{x_{r_i}}^{x_{r_{i+1}}} f(x)\;dx\;.$$

Hence, $f(x_{r_1},\;x_{r_2},\ldots,\;x_{r_k})\;\;dx_{r_1}\;dx_{r_2}\;\ldots\;dx_{r_k}$

$$= \frac{n!}{(r_1-1)!\,(r_2-r_1-1)!\ldots(n-r_k)!}\left[\int_{-\infty}^{x_{r_1}} f(x)\,dx\right]^{r_1-1}\left[\int_{x_{r_1}}^{x_{r_2}} f(x)\,dx\right]^{(r_2-r_1-1)} \cdots \left[\int_{x_{r_k}}^{\infty} f(x)\,dx\right]^{n-r_k}$$

$$f(x_{r_1})\,dx_{r_1}\quad f(x_{r_2})\,dx_{r_2}\quad\ldots\quad f(x_{r_k})\,dx_{r_k}.$$

<u>Distribution of the largest and smallest values in a sample</u>. To apply the same technique to the joint distribution of the largest and smallest values of x in a sample, 5 intervals have been taken, i.e., $2K+1 = 5$. Since $r_1 = 1$ and $r_k = r_2 = n$, it is necessary to consider the probability of obtaining 0 elements in I_1, 1 element in I_2, n-2 elements in I_3, 1 element in I_4, and 0 elements in I_5. From the preceding discussion, it follows that:

$$f(x_1,\, x_n)\,dx_1\,dx_n = \frac{n!}{0!1!(n-2)!\,1!\,0!}$$

$$\left[\int_{-\infty}^{x_1} f(x)\,dx\right]^{0}\left[\int_{x_1}^{x_1+dx_1} f(x)\,dx\right]\left[\int_{x_1+dx_1}^{x_n} f(x)\,dx\right]^{n-2}$$

$$\left[\int_{x_n}^{x_n+dx_n} f(x)\,dx\right]\left[\int_{x_n+dx_n}^{\infty} f(x)\,dx\right]^{0}.$$

The making of the substitutions, as before, gives

$$f(x_1, x_n)\, dx_1 dx_n = n\,(n-1) \left[\int_{x_1}^{x_n} f(x)dx\right]^{n-2} f(x_1)\, dx_1\; f(x_n)\, dx_n.$$

Distribution of range. Wilks[10] then obtained the distribution of the sample range by letting

$$x_n - x_1 = R$$
$$x_1 = S$$

and integrating the resulting distribution with respect to S. He illustrated this method by means of the rectangular distribution.

$$f(x) = 1/r \qquad 0 < x < r$$
$$= 0, \text{ otherwise.}$$

This problem is perfectly straightforward, but difficulty is encountered when the distribution is of a less simple nature.

II. MOMENTS OF RANGE

Tippett[11] says that the distribution of range cannot

---

[10] Ibid., p. 92.

[11] L. H. C. Tippett, "On the Extreme Individuals and the Range of Samples Taken from a Normal Population," Biometrika, XVII (1925) Parts 3 and 4, 368.

be written in any useful form, that the problem has been to find the first four moments so that an appropriate Pearson curve can be fitted, adding that such curves fit actual data sufficiently well to establish the adequacy of the method for practical purposes.

Deriving the equations. The first method of finding the mean range considered by Tippett involves the use of Karl Pearson's[12] expression for the mean difference between the p^{th} and $(p+1)$th individual, the expression having been obtained in the following manner:

The frequency distribution was represented by $y = N\phi(x)$, with no hypothesis as to the nature of the distribution. N was the number of individuals; A, the area to the left of any ordinate y at a character value x; $N - A$, the area to the right. Then if $\alpha = A/N$

$$\alpha = \int_{-\infty}^{x} \phi(x)\, dx.$$

The chance of any random individual having a character value less than x is $A/N = \alpha$ and the chance of having a character value greater than x is $(N - A)/N = 1 - \alpha$.

x_p corresponded to the p^{th} individual's character, and x_{p+1} to the next or $(p+1)$th individual's character. The

---

[12] Karl Pearson, "Note on Francis Galton's Problem," Biometrika, I (1902), 391-92.

problem was to find the mean value $x_p - x_{p+1}$, there being $p - 1$ individuals to the right of y_p and $n - p - 1$ to the left of y_{p+1} in a sample of size n. The chance of an individual falling at x_p is given by $y_p dx_p/N$, and one at x_{p+1}, by $y_{p+1}\, dx_{p+1}/N$; the chance of an individual to the left of y_{p+1} is A_{p+1}/N and to the right of Y_p is $(N - A_p)/N$. The total chance of this combination is given by

$$\frac{y_p\, dx_p}{N} \cdot \frac{Y_{p+1}\, dx_{p+1}}{N} \cdot \left[\frac{A_{p+1}}{N}\right]^{n-p-1} \left[\frac{N - A_p}{N}\right]^{p-1} .$$

But the two individuals can be permuted in $n!/\,(n-p-1)!\,(p-1)!$ ways. To get the average Pearson multiplied the chance thus obtained by the corresponding $x_p - x_{p+1}$, and integrated from $x_{p+1} = -\infty$ to x_p and for $x_p = -\infty$ to ∞, writing x' for x_{p+1}, x for x_p, α' for A_{p+1}/N, α for A_p/N, y_o' for y_{p+1}/N, y_o for y_p/N, thus obtaining for X_p, the average interval between the p^{th} and the $(p+1)$ individual.

$$X_p = \frac{n!}{(n-p-1)!(p-1)!} \int_{-\infty}^{\infty} dx \int_{-\infty}^{x} dx' \;\; y_o \cdot y_o' \, \alpha'^{\,n-p-1} (1-\alpha)^{p-1} (x-x')$$

where

$$\frac{d\alpha'}{dx'} = y_o' \; , \quad \frac{d\alpha}{dx} = y_o \, .$$

Now the x' integral was considered.

$$I = \int_{-\infty}^{x} y_o' \, \alpha'^{\,n-p-1} \, (x-x') \, dx'$$

$$= \int_{-\infty}^{x} \alpha'^{\,n-p-1} \, (x-x') \, d\alpha' .$$

Integration by parts gave

$$\left[\frac{\alpha'^{\,n-p}}{n-p} \, (x-x') \right]_{-\infty}^{x} + \int_{-\infty}^{x} \frac{\alpha'^{\,n-p}}{n-p} \, dx'$$

or between the limits

$$\frac{1}{n-p} \int_{-\infty}^{x} \alpha'^{\,n-p} \, dx' = \frac{1}{n-p} U, \text{ say.}$$

Thus

$$X_p = \frac{n!}{(n-p)!\,(p-1)!} \int_{-\infty}^{\infty} y_o \, U \, (1-\alpha)^{p-1} \, dx$$

$$= \frac{n!}{(n-p)!\,(p-1)!} \int_{-\infty}^{\infty} U \, (1-\alpha)^{p-1} \, d\alpha$$

$$= \frac{n!}{(n-p)!\,p!} \left\{ \left[-U \, (1-\alpha)^{p} \right]_{-\infty}^{\infty} + \int_{-\infty}^{\infty} \frac{dU}{dx} \, (1-\alpha)^{p} dx \right\}$$

or taking the limits and substituting $\frac{dU}{dx}$

$$x_p = \frac{n!}{(n-p)!\ p!} \int_{-\infty}^{\infty} \alpha^{n-p} (1-\alpha)^p \, dx .$$

This then was the formula used by Tippett. The sum of these mean differences for all values of p from 1 to n-1 gives the mean range, $\bar{w}$.

$$\bar{w} = \int_{-\infty}^{\infty} \left[1 - (1-\alpha)^n - \alpha^n \right] dx.$$

Tippett[13] also gave a second method of finding the mean range. Figure 1 represents the curve of the distribution of the original population, $y = \phi(x)$, and as before

$$\alpha_p = \int_{-\infty}^{x_p} \phi(x)\, dx \text{ , where } \int_{-\infty}^{\infty} \phi(x)\, dx = 1.$$

If x_1 is the character of the first individual, and x_n that of the last in a sample of size n, then, on the assumption that the original population is infinite, the chance that there is one individual at x_1, one at x_n, and n-2 between is given by

[13] L. H. C. Tippett, "On the Extreme Individuals and the Range of Samples Taken from a Normal Population," 368-70.

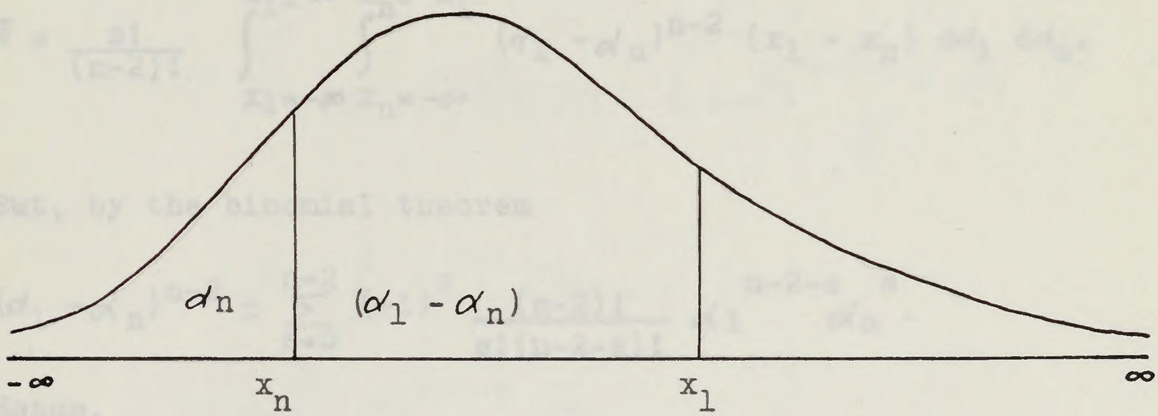

FIGURE 1

CURVE OF DISTRIBUTION, $y = \phi(x)$

$$\frac{n!}{(n-2)!} \quad (\alpha_1 - \alpha_n)^{n-2} \, d\alpha_1 \, d\alpha_n.$$

This expression is equivalent to that of Wilks, except for the fact that Wilks arranged his individuals in ascending order of magnitude; Tippett, in descending. The notion of expected value leads to the defining of the mean range as

$$\bar{w} = \frac{n!}{(n-2)!} \int_{x_1=-\infty}^{x_1=\infty} \int_{x_n=-\infty}^{x_n=x_1} (\alpha_1 - \alpha_n)^{n-2} (x_1 - x_n) \, d\alpha_1 \, d\alpha_n.$$

But, by the binomial theorem

$$(\alpha_1 - \alpha_n)^{n-2} = \sum_{s=0}^{n-2} (-1)^s \frac{(n-2)!}{s!(n-2-s)!} \alpha_1^{\,n-2-s} \alpha_n^{\,s}.$$

Hence,

$$\bar{w} = n! \sum_{s=0}^{n-2} \frac{(-1)^s}{s!(n-2-s)!} \int_{x_1=-\infty}^{x_1=\infty} \alpha_1^{\,n-2-s} \, d\alpha_1 \int_{x_n=-\infty}^{x_n=x_1} \alpha_n^{\,s} (x_1 - x_n) \, d\alpha_n.$$

Now

$$\int_{x_n=-\infty}^{x_n=x_1} \alpha_n^{\,s} (x_1 - x_n) \, d\alpha_n = \left[\frac{(x_1 - x_n)\alpha_n^{s+1}}{s+1} \right]_{x_n=-\infty}^{x_n=x_1}$$

$$+ \frac{1}{s+1} \int_{x_n=-\infty}^{x_n=x_1} \alpha_n^{\,s+1} \, dx_n.$$

The term in brackets vanishes at both limits, since

$$x_1 - x_n = 0, \qquad \text{when} \quad x_n = x_1,$$

and

$$\alpha_n^{s+1} = \left[\int_{-\infty}^{x_n} \phi(x)\, dx\right]^{s+1} = 0, \quad \text{when} \quad x = -\infty,$$

so that, by substitution,

$$\overline{w} = n! \sum_{s=0}^{n-2} \frac{(-1)^s}{(s+1)!\,(n-2-s)!} \int_{x_1=-\infty}^{x_1=\infty} \alpha_1^{\,n-2-s}\, U\, d\alpha_1,$$

where

$$U = \int_{-\infty}^{x_1} \alpha_n^{\,s+1}\, dx_n.$$

If

$$\theta = \int_{\alpha_1}^{1} \alpha_1^{\,n-2-s}\, d\alpha_1$$

$$= \frac{1-\alpha_1^{\,n-1-s}}{n-1-s},$$

then

$$\frac{d\theta}{d\alpha_1} = -\alpha_1^{\,n-2-s}$$

and $$\bar{w} = - n! \sum_{s=o}^{n-2} \frac{(-1)^s}{(s+1)!\,(n-2-s)!} \int_{x_1=-\infty}^{x_1=\infty} U \frac{d\theta}{d\alpha_1} \, d\alpha_1 .$$

Integration by parts gives

$$\bar{w} = n! \sum_{s=o}^{n-2} \frac{(-1)^s}{(s+1)!\,(n-2-s)!} \left\{ - \Big[U \,\theta \Big]_{x_1=-\infty}^{x_1=\infty} + \int_{-\infty}^{\infty} \theta \frac{dU}{dx_1} \, dx_1 \right\}.$$

Since the term in brackets vanishes and

$$\frac{dU}{dx_1} = \frac{d}{dx_1} \int_{-\infty}^{x_1} \alpha_n^{\,s+1} \, dx_n = \frac{d}{dx_1} \int_{-\infty}^{x_1} \left[\int_{-\infty}^{x_n} \phi(x)\,dx \right]^{(s+1)} dx_n$$

$$= \alpha_1^{\,s+1}$$

$$\bar{w} = \sum_{s=o}^{n-2} (-1)^s \frac{n!}{(s+1)!\,(n-1-s)!} \int_{-\infty}^{\infty} (1-\alpha_1)^{n-1-s} \alpha_1^{\,s+1} \, dx,$$

which leads to Tippett's previous result

$$\bar{w} = \int_{-\infty}^{\infty} \Big[1 - (1-\alpha)^n - \alpha^n \Big] \, dx.$$

The higher moments may be obtained in a similar manner so that for even values of n

$$\mu_m = m\,(m-1) \int_{-\infty}^{\infty}\int_{-\infty}^{x_1} \left[1 - \alpha_1^{\,n} - (1-\alpha_n)^n + (1-\alpha_n)^n\right]$$

$$(x_1 - x_n - \bar{w})^{m-2} \; dx_1 \, dx_n - (m-1)\,(-\bar{w})^m.$$

On putting m = 2,

$$\mu_2 = 2 \int_{-\infty}^{\infty}\int_{-\infty}^{x_1} \left[1 - \alpha_1^{\,n} - (1 - \alpha_n)^n\right.$$

$$\left. + (1 - \alpha_n)^n\right] \; dx_1 \; dx_n - \bar{w}^2 \;.$$

Constructing the tables. From the equation for the mean range when the samples are from a normal population Tippett[14] found a framework of values by direct computation, using quadratures. This he filled in by interpolation using first Lagrangian Formulae; and lastly, a difference formula. The result was his table for the mean range for a normal distribution for samples of size n from 2 to 1000. The values are for a population having unit standard deviation,

[14] Ibid., pp. 371-73.

so that to obtain the absolute range in any given case, the tabled value must be multiplied by the actual value of the standard deviation. Figure 2 illustrates Tippett's results graphically.

A similar method was used in the case of the second moment and the standard deviation. The framework was constructed by substitution in the formula. The process involving cubature was very laborious. His results have been shown in Figure 3.

Much time was spent in trying to evaluate the third and fourth moments by this same method, but many difficulties were encountered and the results obtained were irregular. One cause of difficulty was the fact that the equation consists of two nearly equal parts, one of which must be subtracted from the other, so that the computations must be very accurate if the difference is to be relied upon. Consequently, Tippett resorted to a method of obtaining μ_3 and μ_4 from the separate distributions of the first and last individuals. He started with the following general formulas:

$$\bar{w} = \bar{u} - \bar{v}$$
$$\mu_{2_w} = 2\mu_{2_u}(1-r)$$
$$\mu_{3_w} = 2\mu_{3_u} + 6p_{12}$$
$$\mu_{4_w} = 2\mu_{4_u} - 8p_{13} + 6p_{22},$$

where the p's are certain product moment coefficients.

so that to obtain the absolute range in any given case, the tabled value must be multiplied by the actual value of the standard deviation. Figure 2 illustrates Tippett's results graphically.

A similar method was used in the case of the second moment and the standard deviation. The framework was constructed by substitution in the formula. The process involving cubature was very laborious. His results have been shown in Figure 3.

Much time was spent in trying to evaluate the third and fourth moments by this same method, but many difficulties were encountered and the results obtained were irregular. One cause of difficulty was the fact that the equation consists of two nearly equal parts, one of which must be subtracted from the other, so that the computations must be very accurate if the difference is to be relied upon. Consequently, Tippett resorted to a method of obtaining μ_3 and μ_4 from the separate distributions of the first and last individuals. He started with the following general formulas:

$$\bar{w} = \bar{u} - \bar{v}$$
$$\mu_{2w} = 2\mu_{2u}(1-r)$$
$$\mu_{3w} = 2\mu_{3u} + 6p_{12}$$
$$\mu_{4w} = 2\mu_{4u} - 8p_{13} + 6p_{22},$$

where the p's are certain product moment coefficients.

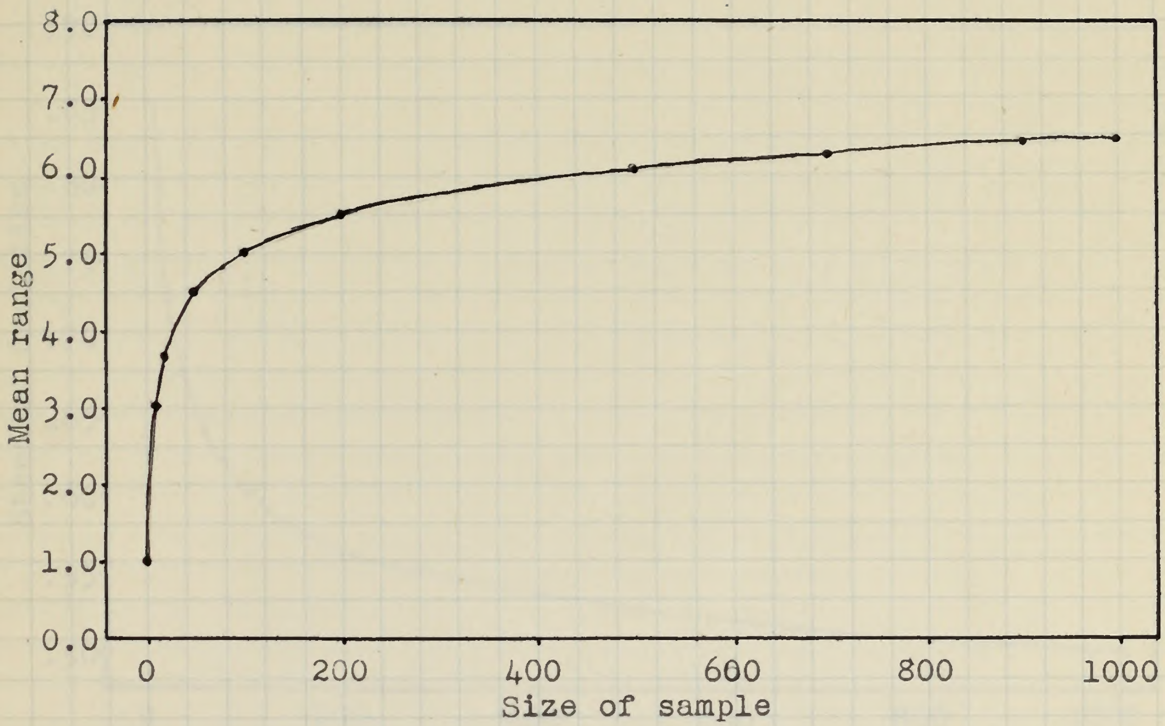

FIGURE 2

MEAN RANGE IN SAMPLES OF SIZE n (= 2 to 1000) IN TERMS OF STANDARD DEVIATION OF SAMPLED POPULATION.

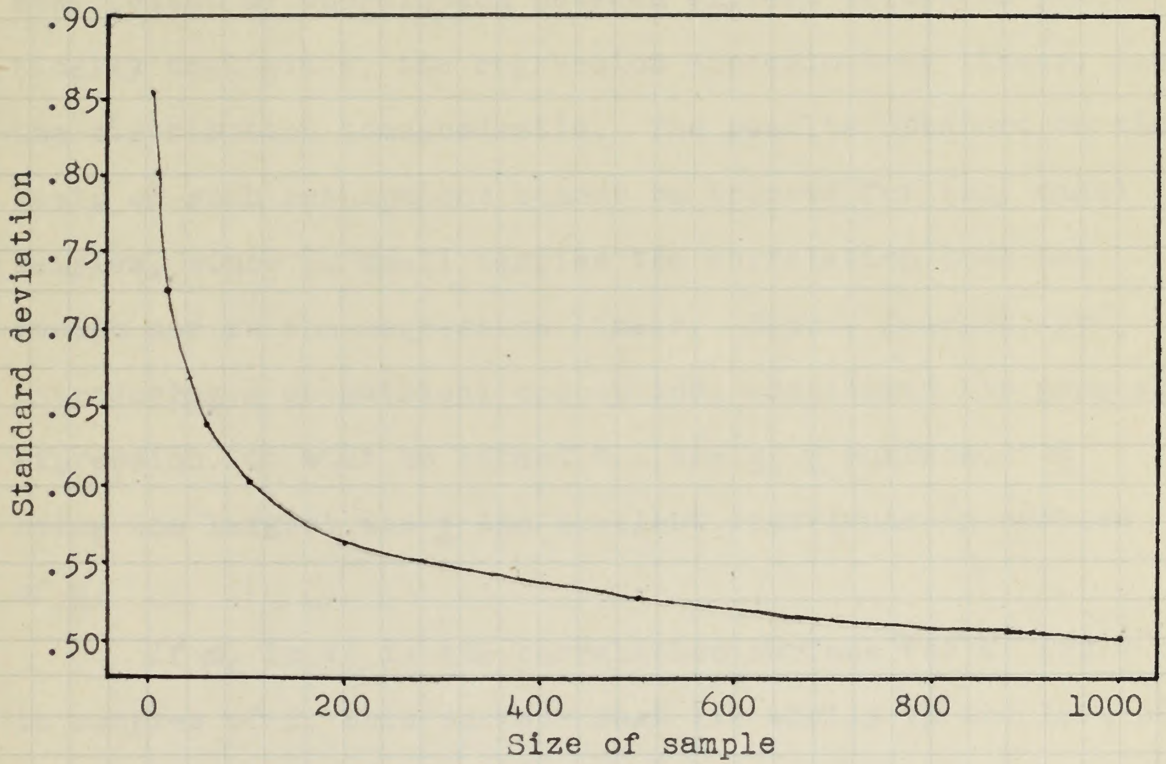

FIGURE 3

STANDARD DEVIATION OF RANGE IN SAMPLES OF SIZE n (= 2 to 1000) IN TERMS OF STANDARD DEVIATION OF SAMPLED POPULATION.

E. S. Pearson,[15] however, called attention to the fact that Tippett, in obtaining the moments, simplified these formulas on the assumption that for samples of size 60 or more the coefficient of correlation between extreme values is practically negligible, the regression approximately linear, and the distribution homoscedastic. The results obtained on the basis of such assumptions cannot be trusted for very small samples, since in small samples the correlation does not vanish nor is the regression linear. Hence, Pearson, by introducing a geometrical conception, considered the general expression for what he termed "...the _u_, _v_ surface," _u_ being the largest and _v_ the smallest individual in samples of _n_.

If $\varphi_n\,(u,v)$ is the correlation surface for u and v in samples of n, this surface must lie wholly to the left of $u = v$, since $u \geq v$. Now if samples of $n+1$ are taken, the surface $\varphi_n\,(U,V)$ differs from $\varphi_n\,(u,v)$, since x, the $n+1$ th individual, may be such that $u > x > v$, or $u > v > x$, or $x > u > v$. If the sampling is from a normal population with unit standard deviation, the distribution of this single individual is

[15] E. S. Pearson, "A Further Note on the Distribution of Range in Samples Taken from a Normal Population," _Biometrika_, XVII (1926), 173-93.

$$\frac{1}{\sqrt{2\pi}} \quad e^{-\frac{1}{2}x^2}.$$

Pearson considered the frequency distribution of x in each case and by summing the three contributions corresponding to every value of x possible he obtained the complete frequency surface $\phi_{n+1}(U, V)$:

$$\phi_{n+1}(U, V) = \phi_n(U, V) \int_V^U \frac{e^{-\frac{1}{2}x^2}}{\sqrt{2\pi}} dx + \frac{e^{-\frac{1}{2}V^2}}{\sqrt{2\pi}} \int_V^U \phi_n(U, v)\, dv$$

$$+ \frac{e^{-\frac{1}{2}V^2}}{\sqrt{2\pi}} \int_V^U \phi_n(u, V)\, du.$$

This provided a reduction formula for obtaining the correlation surface of the extreme individuals in samples of n in terms of the equations of surfaces for smaller samples. By further substitutions this formula was reduced to

$$z = \frac{n\,(n-1)}{2\pi}\; e^{-\frac{1}{2}(u^2+v^2)}\; (A_u - A_v)^{n-2}$$

where $A_t = \int_{-\infty}^{t} \frac{e^{-\frac{1}{2}x^2}}{\sqrt{2\pi}}\, dx.$

This equation together with that for the frequency distribution of u, given by

$$\phi(u) = \frac{n}{\sqrt{2\pi}} \; e^{-\frac{1}{2}u^2} \; A_u^{\;n-1}$$

when the sample is drawn from a normal population, was used to find the moments and product moment coefficients needed for Tippett's general relations for the moments of range. The expressions involved cannot in general be integrated, but Pearson completed the solution for the cases of $n = 2, 3, 4, 5,$ and 6, by using integrals which he evaluated by quadrature. He found that his values for the mean ranges corresponded exactly with those obtained by Tippett. The values he obtained for σ_w, are shown in Figure 4, which represents on an enlarged scale the start of Tippett's curve illustrated in Figure 3.

Pearson further investigated what justification there was for assuming linear regression and homoscedasticity at $n = 6$ in the case of the constants β_1, and β_2. The differences between the results obtained from the general expression and those based on the preceding assumptions were too great to warrant the use of these assumptions. In fact, he came to the conclusion that, however great n may be, the regression can never be strictly linear, nor can theoretical homoscedasticity be obtained over the whole surface. Still, both assumptions may be justified in the region of significant frequency, and as n increases the coefficient of correlation tends to zero. Hence, for

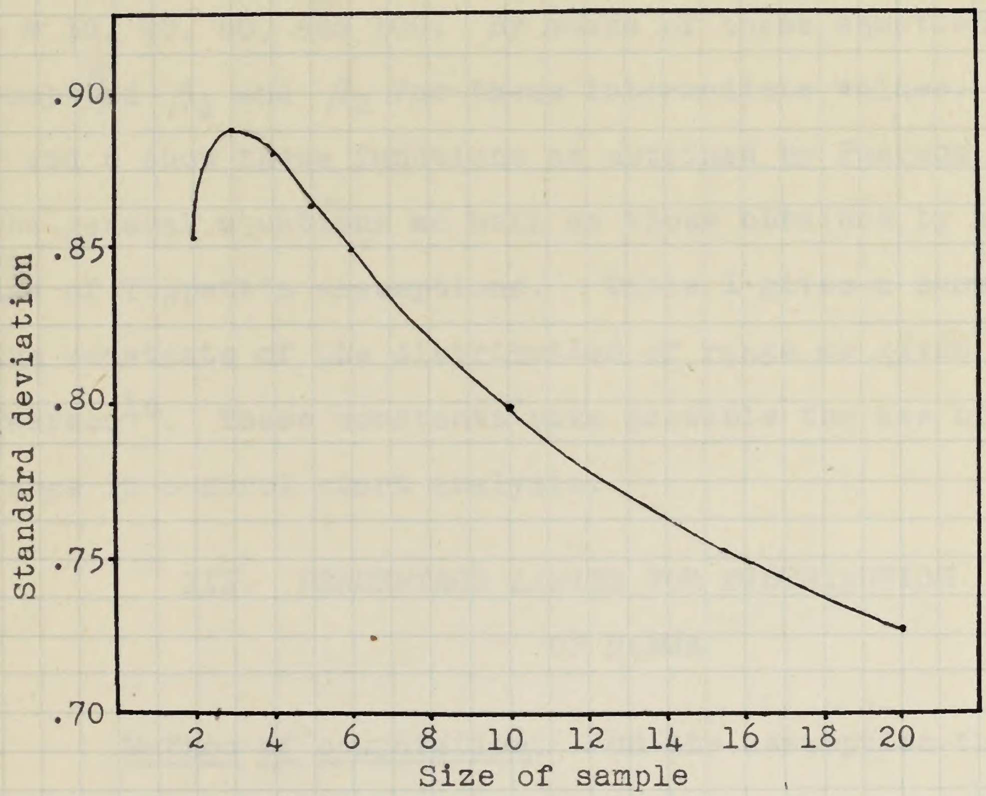

FIGURE 4

STANDARD DEVIATION OF RANGE IN SAMPLES OF SIZE n (= 2 to 20) IN TERMS OF STANDARD DEVIATION OF SAMPLED POPULATION.

samples of 60 or more serious error can hardly be involved in using Tippett's constants β_1 and β_2. To bridge the gap between n = 6 and n = 60, Pearson determined the equations of the "best-fitting" regression parabolae for n = 10, 20, 60, and 100. By means of these equations he computed β_1 and β_2 for these intermediate values. Figures 5 and 6 show these functions as obtained by Pearson from the general equations as well as those obtained by making use of Tippett's assumptions. Table I gives a summary of the constants of the distribution of range as given by Pearson[16]. These constants make possible the use of the range in control chart analysis.

III. PERCENTAGE LIMITS FOR DISTRIBUTION OF RANGE

<u>Method of computation</u>. On the assumption that the distribution of the range may be adequately represented by Pearson curves with appropriate moment coefficients, E. S. Pearson[17] obtained a framework by finding equations of Type I and Type VI curves, using appropriate frequency con-

[16] <u>Ibid</u>., p. 192.

[17] E. S. Pearson, "The Percentage Limits for the Distribution of Range in Samples from a Normal Population," <u>Biometrika</u>, XXIV (1932), 404-7.

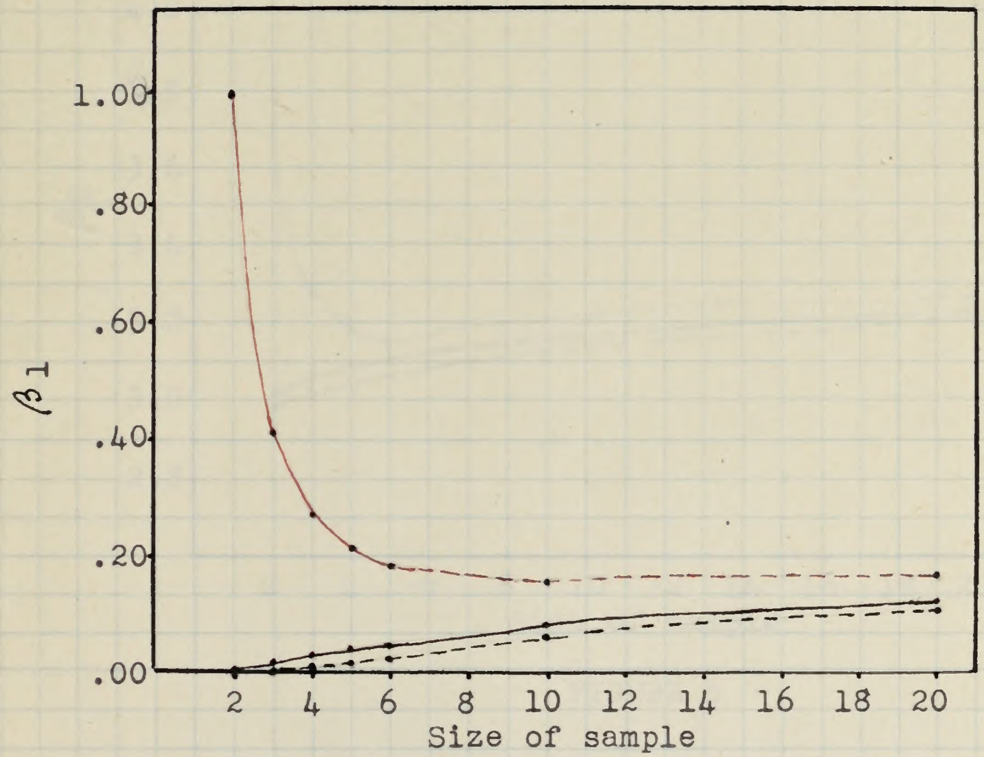

FIGURE 5

β_1 OF RANGE

—— True values
- - - - Approximate continuation
—— Assuming zero correlation and homoscedasticity
- - - - Assuming linear regression and homoscedasticity

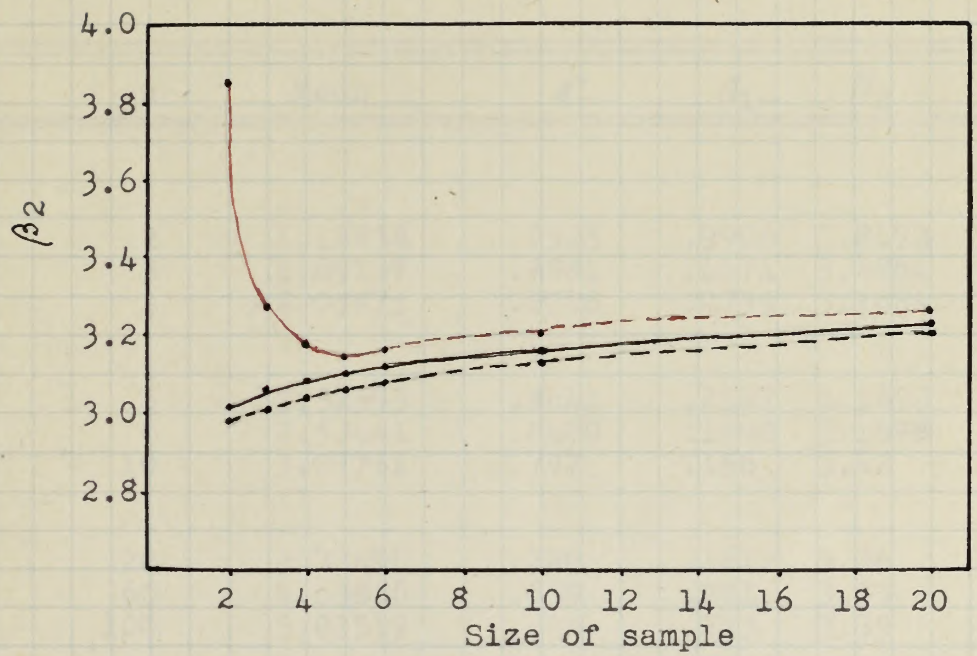

FIGURE 6

β_2 OF RANGE

——— (red)	True values
- - - - (red)	Approximate continuation
———	Assuming zero correlation and homoscedasticity
- - - -	Assuming linear regression and homoscedasticity

TABLE I

SUMMARY OF CONSTANTS OF DISTRIBUTION OF RANGE

n	Mean	σ	β_1	β_2
2	1.12838	.8525	.9906	3.8692
3	1.69257	.8884	.4174	3.2864
4	2.05875	.8798	.2735	3.1884
5	2.32593	.8641	.2167	3.1693
6	2.53441	.8480	.1892	3.1698
10	3.07751	.797	.156	3.22
20	3.73495	.729	.161	3.26
60	4.63856	.639	.201	3.35
100	5.01519	.605	.223	3.39
200	5.49209	.566	.247	3.44
500	6.07340	.524	.285	3.50
1000	6.48287	.497	.309	3.54

stants for samples of size

$n = 3, 4, 6, 10, 20, 60, 100.$

The first four curves were made to start at the point, range = $w = o$, and were given the correct mean ($=\bar{w}$) standard deviation ($=\sigma_w$) and β_1. For the other three curves the start was not fixed, but the first four theoretical moments used — $\bar{w}$, σ_w, β_1, and β_2. The use of different methods is accounted for by the fact that when n is small the distribution of the range is abrupt at the lower end. Hence, it seemed advantageous to give the curves the correct start. On the other hand, as n increased, it seemed advisable to use the correct β_2 rather than the correct start.

The percentage limits computed were the upper and lower 0.5, 1, 5, and 10, thus giving the boundaries within which 99%, 98%, 90%, and 80% of the ranges would lie. The position of the ordinate at the upper and lower limits for each of the framework curves was found by quadrature and backward interpolation. For a given percentage limit, p, the value l_p (the position of the ordinate) changes with n, that is, with β_1 and β_2 or with the shape of the sampling curve. It was found that the change was not rapid so that it was possible to find by interpolation in the framework each of the eight values of l_p for

$n = 3, 4, 5, \ldots\ldots, 29, 30, 35, 40, \ldots\ldots 95, 100.$

Since $l_p = (w_p - \bar{w})/\sigma_w$, where w_p represents the range value corresponding to any one of the ordinates, having calculated the l_p's, it was only necessary to invert the formula to obtain w_p from

$$w_p = \bar{w} + l_p \sigma_w.$$

Pearson used values of $\bar{w}$ as computed by Tippett, but σ_w had only been calculated for
$n = 2, 3, 4, 5, 6, 10, 20, 60, 100.$
Three additional values were computed at

$$n = 30, 45, 75$$

by the same process of cubature as that employed by Tippett. From this framework the intermediate values of σ_w were obtained and finally the values of l_p given in Pearson's Table A.[18] Since the values of $\bar{w}$ and σ_w are for samples drawn from a normal population, this table gives the percentage limits for the distribution of range in samples from a normal population.

In a more recent article Pearson[19] stated that no simple expression exists for the probability law $f_n(w)$ of w, but he gave a table of computed values of the pro-

---

[18] Ibid., p. 416.

[19] E. S. Pearson, "The Probability Integral of the Rangeiin Samples of Observations from a Normal Population," Biometrika, XXXII (1942), Parts 3 and 4, 301-7.

bability integral, which gives the chance that the range in a sample of n observations is less than a given multiple of the population standard deviation. These values are more accurate than those previously obtained by Pearson and from them any confidence limits can be obtained. Hartley[20] showed the derivation of the complicated formula used and described the numerical evaluation of the probability integral which was accomplished by means of quadrature formulas and machine calculation.

IV. CONCLUSION

From this investigation and discussion it has been concluded that the sampling distribution of the range is asymmetrical but approaches most nearly to normal when $6 < n < 10$. The method of moments offered the most satisfactory approach to the study of this distribution, with the moments calculated from the general equations with no assumptions as to linearity of regression or homoscedasticity of the correlation surface. Confidence limits were best obtained by computing the values of the probability integral according to values of the range and the size of the sample. All constants of the distribution of range were given in terms of the standard deviation of the population.

[20] H. O. Hartley, "The Range in Random Samples," Biometrika, XXXII (1942), Parts 3 and 4, 341-42.

PART II

APPLICATION OF RANGE TO QUALITY CONTROL

It has been previously stated that the range finds its principal use in its application to quality control. Specifically it is used in control chart analysis. In this respect the range can be employed to estimate the standard deviation of the population for control chart purposes or the control chart may be made for the ranges instead of for the standard deviations of the sample.

I. QUALITY CONTROL TECHNIQUE

Nature of quality control. The purpose of this paper is not to discuss quality control. Nevertheless, the subject of range cannot be adequately treated without some explanation of it. The idea of control involves action for the purpose of achieving a desired end, for example, detecting causes of trouble in processes, securing conformity with specifications, estimating the quality of a product, or the like. A manufacturer wishes to control a certain quality characteristic in his product. He cannot maintain an entirely uniform product. Yet if each factor which might cause variation in this characteristic continues during the process of manufacture to have the same probability of contributing a given effect, then the particular quality may

be said to be controlled in a technical sense.[21]

If the cost of testing the entire output is prohibitive, or if the test is destructive, it will be necessary to resort to sampling in order to determine whether uniformity of quality is being maintained. This involves statistical control.[22] A standard level must be set and the limits within which the measurements of the quality characteristic may fluctuate without on the average departing from this level must be determined. The statistics for the successive samples must be recorded and comparisons made with the control limits.

The control chart. Now the control chart is a graphic representation of this type of data. On it are pictured the central value, or quality level desired, and the upper and lower control limits, that is, the boundaries within which the measures must remain if a state of control is to be maintained. When the findings have been plotted from sample to sample, the character of the output can be seen at a glance. Whether the quality of the product is unsatisfactory because the level of control does not

[21]Frederick E. Croxton and Dudley J. Cowden, Applied General Statistics (New York: Prentice-Hall, Inc. 1941), p. 348.

[22]Ibid., p. 349.

remain constant, that is changes from one consignment to another, or because the level maintained, though constant, is too low, either because the mean of a variable characteristic is too small or because the standard deviation is too large --- in either case there is a type of control chart to suit the purpose. Changes in the level of control are most likely to be detected by an examination of the control chart for the means. But even when the level of control is maintained, individual units may fall below specification, if the variation within the sample is large. Satisfactory quality as well as a uniform product can be assured by the employment of a control chart for some measure of variability, such as the standard deviation. Other measures, especially the range, may be substituted for the standard deviation in such a chart.

Regardless of the form of chart used the standard deviation of the population is needed to set the control limits. If past experience has provided a standard of quality, for example, the population mean and standard deviation are known, the problem will simply be to discover whether fresh material continues to conform uniformly to this standard. If, on the other hand, no standard has been fixed, the standard deviation of the population must first

be determined.[23]

Methods of determining σ. E. S. Pearson[24] has given three methods of estimating this statistic:

(1) From the mean value of the variance:

$$\sigma_c^2 = \frac{n}{N-k}(s_1^2 + s_2^2 + \cdots + s_k^2),$$

where N is the total number of observations, k the number of subgroups, n the number of units in each subgroup. σ_c will be the square root of the result.

(2) From the mean value of the standard deviation:

$$\sigma_c = \frac{1}{b_n} \cdot \frac{1}{k}(s_1 + s_2 + \cdots + s_k),$$

where the factor, b_n, is the ratio of the mean value of the standard deviation of the samples to the standard deviation of the population for samples of size n. Pearson gives tables for b_n and $1/b_n$ for values of n from 2 to 30.

(3) From the mean value of the range:

$$\sigma_c = \frac{1}{d_n} \cdot \frac{1}{k}(w_1 + w_2 + \cdots + w_k),$$

where d_n is the ratio of the mean range to the standard deviation of the population for samples of size n. Pearson gives

---

[23] E. S. Pearson, The Application of Statistical Methods to Industrial Standardisation and Quality Control (London: British Standards Institution, 1935), p. 82.

[24] Ibid., pp. 83-84.

tables for d_n and $1/d_n$ for values of n from 2 to 12. Tippett[25] has given full tables for this ratio for samples of between 2 and 1000 individuals drawn from a normal population. An abstract of these tables is given in Table II.

This is the table of mean ranges referred to in Part I. Since the mean range in these tables is in terms of the standard deviation of the original population, the latter can, in a particular case, be found by taking samples, determining the mean range, and dividing by the value given in the tables. The method is similar to one given by K. Pearson [26] in which the sample is ranked and the difference between two certain individuals, preferably those near the quindeciles (those n/15 from each end), measured and divided by the value for a population having unit standard deviation.

From Table II it is seen how much the range depends on the size of the sample. Freeman[27] stated that for small k, say less than 10, or large n, say greater than 15, the mean range method of estimating σ is unreliable, but that

---

[25] L. H. C. Tippett, "On the Extreme Individuals and the range of Samples Taken from a Normal Population," Biometrika, XVII (1925), Parts 3 and 4, p. 386.

[26] K. Pearson, (Editorial) "On the Probable Errors of Frequency Constants," Biometrika, XIII (1921), 113-119.

[27] H. A. Freeman, Industrial Statistics (New York: John Wiley & Sons, Inc., 1942), p. 131.

TABLE II

RATIO OF MEAN RANGE TO STANDARD DEVIATION

No. in Sample	$\frac{\text{Mean Range}}{\text{Standard Deviation}}$
2	1.128
4	2.059
5	2.326
10	3.078
50	4.498
100	5.015
500	6.073

for n less than 15 a good estimate of σ can be formed from the mean range. Pearson[28] said that no estimate of σ can be considered satisfactory if the total number of observations is less than 30, but that if N (= nk) is greater than 50 and the observations have been broken up into equal sub-groups each containing not more than 10 units, the estimate will be adequate for control chart purposes. Tippett[29] recommended collecting the data into groups of 5 or 10 units.

II. APPLICATION OF THE TECHNIQUE

The sample. A problem has been considered as best illustrating the approximations of σ obtained by the three methods suggested. Shewhart[30] has given a set of 204 observations of the measurements in megohms of the insulation resistances of as many different pieces of a new kind of material produced under presumably the same essential conditions. The particular characteristic measured was not in a state of control. Nevertheless, the sample will serve for

---

[28] E. S. Pearson, The Application of Statistical Methods, p. 84.

[29] L. H. C. Tippett, The Methods of Statistics (second edition, revised; London: Williams and Norgate, Ltd., 1937) p. 32.

[30] Walter A. Shewhart, Statistical Method from the Viewpoint of Quality Control (Washington: U. S. Department of Agriculture, 1939), p. 90.

illustrative purposes. Table III gives these measurements in the order in which they were taken.

Statistics of the sample. The sample has been divided into 51 subgroups of 4 units each, the order in which the measurements were taken being preserved, since that is the order which should furnish the clue to assignable causes of variability. The means ($\bar{x}$), variances (s^2), standard deviations (s), and ranges (w) of these subgroups are given in Table IV.

Computation of σ. Substitution of the values from Table IV in the formulas above gives the following results:

For N = 204, n = 4, k = 51

From s_i^2 $\quad \sigma_c = \sqrt{\frac{4}{153} \cdot 4834406} = 355.5$

From s_i $\quad \sigma_c = 1.253 \cdot \frac{1}{51} \cdot 13358 = 328.2$

From w_i $\quad \sigma_c = 0.4857 \cdot \frac{1}{51} \cdot 33600 = 320.0$

Control limits. If the control chart is made for the means, the estimate from the mean range is sufficiently accurate and a distinct time-saver. If, on the other hand, the control chart is made for the standard deviations, the estimate from the mean standard deviation or from the mean variance is more in order, so as to eliminate any unnecessary

TABLE III

OBSERVATIONS OF 204 MEASUREMENTS OF

INSULATION RESISTANCE

5045	4790	4090	5000	4840	5000	4625
4350	4845	5000	4575	4310	4700	4425
4350	4700	4335	4700	4185	4500	4135
3975	4600	5000	4430	4570	4840	4190
4290	4100	4640	4850	4700	5075	4080
4430	4410	4335	4850	4440	5000	3690
4485	4180	5000	4570	4850	4770	5050
4285	4790	4615	4570	4125	4570	4625
3980	4790	4215	4855	4450	4925	5150
3925	4340	4275	4160	4450	4775	5250
3645	4895	4275	4325	4850	5075	5000
3760	5750	5000	4125	4450	4925	5000
3300	4740	4615	4100	3635	5075	
3685	5000	4735	4340	3635	4925	
3463	4895	4215	4575	3635	5250	
5200	4255	4700	3875	3900	4915	
5100	4170	4700	4050	4340	5600	
4635	3850	4700	4050	4340	5075	
5100	4445	4700	4685	3665	4450	
5450	4650	4095	4685	3775	4215	
4635	4170	4095	4430	5000	4325	
4720	4255	3940	4300	4850	4665	
4810	4170	3700	4690	4775	4615	
4565	4375	3650	4560	4500	4615	
4410	4175	4445	3075	4770	4500	
4065	4550	4000	2965	4500	4765	
4565	4450	4845	4080	4770	4500	
5190	2855	5000	4080	5150	4500	
4725	2920	4560	4425	4850	4850	
4640	4375	4700	4300	4700	4930	
4640	4375	4310	4430	5000	4700	
4895	4355	4310	4840	5000	4890	

TABLE IV

STATISTICAL MEASURES FOR DISTRIBUTION OF INSULATION RESISTANCE MEASUREMENTS DIVIDED INTO 51 SUBGROUPS OF 4 UNITS EACH

Sample No.	$\bar{\bar{X}}$	s^2	s	w
1	4430	149512	387	1070
2	4372	7606	87	200
3	3828	17656	133	335
4	3912	571654	756	1900
5	5071	83855	290	815
6	4682	8432	92	245
7	4558	166106	408	1125
8	4725	10838	104	255
9	4734	8642	93	245
10	4370	71750	268	690
11	4944	260142	510	1410
12	4722	81406	285	745
13	4279	90280	300	800
14	4242	7056	84	205
15	4008	461606	679	1695
16	4006	393380	627	1455
17	4606	162542	403	910
18	4648	55756	236	665
19	4441	104667	324	785
20	4566	43030	207	520
21	4549	68630	262	605
22	3846	32642	181	445
23	4572	150256	388	1000
24	4470	28050	167	390
25	4676	44067	210	570

Sample No.	$\overline{x}$	s^2	s	w
26	4710	19600	140	280
27	4366	85330	292	730
28	4222	68456	262	700
29	4368	100806	317	635
30	4495	21125	145	390
31	3550	282412	531	1115
32	4499	41530	204	540
33	4476	63392	252	655
34	4529	75855	275	725
35	4550	30000	173	400
36	3701	13167	115	265
37	4030	97612	312	675
38	4781	32930	181	500
39	4798	53569	231	650
40	4888	15469	124	300
41	4760	33800	184	500
42	4854	39467	199	505
43	4925	11250	106	300
44	5041	18542	136	335
45	4835	293862	542	1385
46	4555	18050	134	340
47	4566	13167	115	265
48	4842	7569	87	230
49	4344	38230	196	490
50	4361	268405	518	1360
51	5100	11250	106	250
Totals		4834406	13358	33600

computations. But here is an important point. Because of this relation between the range and the standard deviation a control chart made for the range exhibits the same variations as the control chart for the standard deviations. Plotting the data of Table IV for standard deviation and for range, as in Figures 7 and 8, illustrated this fact.

E. S. Pearson[31] has fixed the control limits for the standard deviation as follows:

Outer control limits at $B_{0.001} \cdot \sigma$ and $B_{0.999} \cdot \sigma$

Inner control limits at $B_{0.025} \cdot \sigma$ and $B_{0.975} \cdot \sigma$

Mean value of $s = b_n \sigma$.

The quantities B and b for samples of from 2 to 30 units are given in his text and are for $n = 4$: $B_{0.001} = 0.078$, $B_{0.025} = 0.232$, $B_{0.975} = 1.529$, $B_{0.001} = 2.017$.[32] These values give an outer pair of limits within which (were the variation statistically uniform) 99.8 per cent of the values of s_i should fall and an inner pair within which 95 per cent should fall.

For the range the 95 per cent and 99.8 per cent limits are set up in a similar manner.

---

[31] E. S. Pearson, _The Application of Statistical Methods_, p. 88.

[32] _Ibid._, p.86.

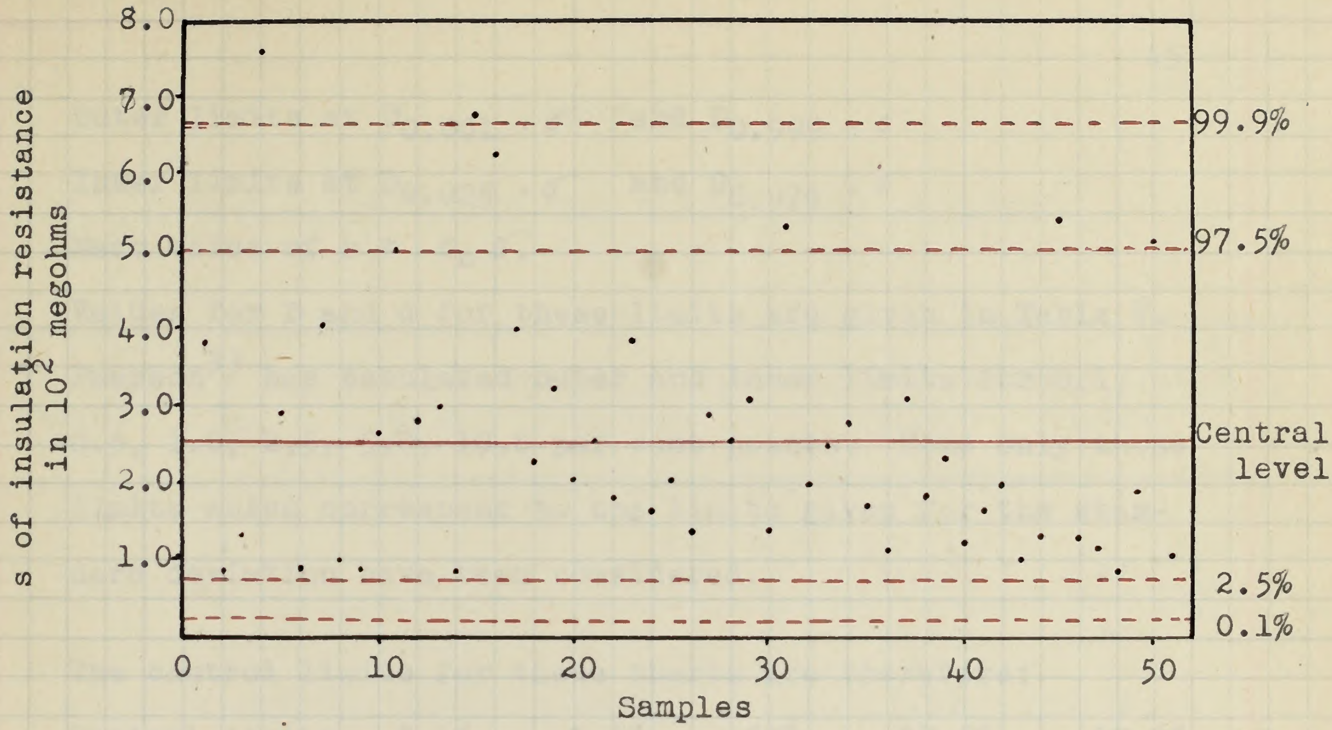

FIGURE 7

CONTROL CHART FOR STANDARD DEVIATION

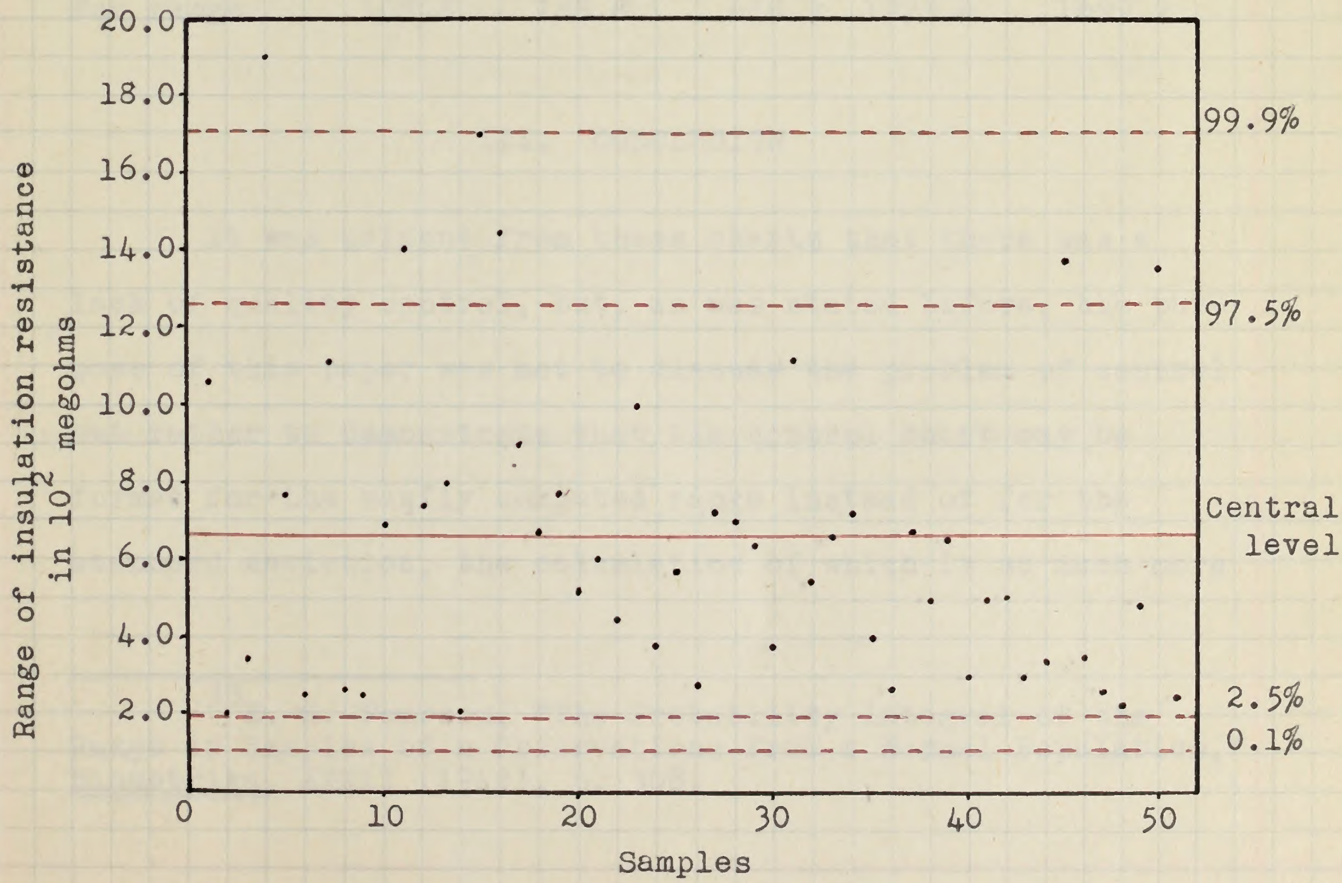

FIGURE 8

CONTROL CHART FOR RANGE

Outer limits at $D_{0.001} \cdot \sigma$ and $D_{0.999} \cdot \sigma$

Inner limits at $D_{0.025} \cdot \sigma$ and $D_{0.975} \cdot \sigma$

Mean value of $r = d_n \sigma$.

Values for D and d for these limits are given in Table V. Pearson[33] has tabulated outer and inner limits for 0.1, 0.5, 1.0, 2.5, 5.0, 10.0 per cent points. Here only those limits which correspond to the limits given for the standard deviation have been considered.

The control limits for these charts are therefore:

Control limits.	0.1%	2.5%	50%	97.5%	99.9%
For standard deviation	25.6	76.1	261.9	501.8	662.0
For range	108.8	188.8	658.9	1273.6	1699.2

III. CONCLUSION

It was evident from these charts that there was a lack of quality control, but, as was stated before, the purpose of this paper was not to discuss the problem of control but rather to demonstrate that the control chart may be formed for the easily computed range instead of for the standard deviation, the calculation of which is so much more

---

[33] E. S. Pearson, "The Probability Integral of the Range in Samples of n Observations from a Normal Population," Biometrika, XXXII (1942), p. 308.

TABLE V

PROBABILITY LIMITS FOR DISTRIBUTION OF RANGE

Size of Sample	d_n	Lower percentage points		Upper percentage points	
		$D_{0.001}$	$D_{0.025}$	$D_{0.975}$	$D_{0.999}$
2	0.8862	0.00	0.04	3.17	4.65
3	0.5908	0.06	0.30	3.68	5.06
4	0.4857	0.20	0.59	3.98	5.31
5	0.4299	0.37	0.85	4.20	5.48
6	0.3946	0.54	1.06	4.36	5.62
7	0.3698	0.69	1.25	4.49	5.73
8	0.3512	0.83	1.41	4.61	5.82
9	0.3367	0.96	1.55	4.70	5.90
10	0.3249	1.08	1.67	4.79	5.97
11	0.3152	1.20	1.78	4.86	6.04
12	0.3069	1.30	1.88	4.92	6.09

laborious, and that such a chart is not less meaningful. A study of the charts revealed the similarity in pattern between the two, which indicates that the simpler chart may be usefully employed for control purposes. This conclusion has been confirmed by theoretical investigation.

However, Pearson called attention to the following facts regarding the range:

> (1) Apart from these special applications when dealing with a number of small groups, the use of the range is to be deprecated since it provides a far less accurate measure of variation than that given by the standard deviation.
>
> (2) Its use in control chart analysis can only be recommended where each sample contains not more than 10 units, since as n increases the range becomes a less and less reliable measure of variation, depending only on the extreme values and taking no account of the form of variation between these.
>
> (3) The condition that the variation due to chance causes should be of the normal form is somewhat more stringent than in the case of charts for standard deviation.[34]

Nevertheless, in spite of these apparent handicaps, the range has been proven theoretically to be a very useful measure. To some extent it has already been put to practical use and promises in future, once its nature is more thoroughly understood, to have even wider application.

---

[34] E. S. Pearson, The Application of Statistical Methods, p.89.

BIBLIOGRAPHY

Croxton, Frederick E. and Dudley J. Cowden, Applied General Statistics. New York: Prentice-Hall, Inc. 1941. 944 pp.

Fisher, R. A. and L. H. C. Tippett, "Limiting Forms of the Frequency Distribution of the Largest and Smallest Members of a Sample," Proceedings of the Cambridge Philosophical Society, XXIV (1928), Part 2, 180-90.

Freeman, H. A., Industrial Statistics. New York: John Wiley & Sons, Inc., 1942. 178 pp.

Gumbel, E. J., "Ranges and Midranges," Annals of Mathematical Statistics, XV (1944), 414-22.

Hartley, H. O., "Numerical Evaluation of the Probability Integral," Biometrika, XXXII (1942) Parts 3 and 4, 309-10.

_______, "The Range in Random Samples," Biometrika, XXXII (1942), Parts 3 and 4, 334-48.

Pearson, E. S., "A Further Note on the Distribution of Range in Samples Taken from a Normal Population," Biometrika, XVIII (1926), 173-94.

_______, The Application of Statistical Methods to Industrial Standardisation and Quality Control. London: British Standards Institution, 1935. 161 pp.

_______, "The Percentage Limits for the Distribution of Range in Samples from a Normal Population," Biometrika, XXIV (1932), 404-17.

_______, "The Probability Integral of the Range in Samples of n Observations from a Normal Population," Biometrika, XXXII (1942) Parts 3 and 4, 301-8.

Pearson, Karl, "Note on Francis Galton's Problem," Biometrika I (1902), 390-99.

_______, (Editorial) "On the Probable Errors of Frequency Constants." Biometrika, XIII (1921), 113-32.

Richardson, C. H., An Introduction to Statistical Analysis, Enlarged edition; New York: Harcourt, Brace and Company 1941. 312 pp.

Shewhart, Walter A., Statistical Method from the Viewpoint of Quality Control. Washington: U. S. Department of Agriculture, 1939. 155 pp.

Tippett, L. H. C., "On the Extreme Individuals and the Range of Samples Taken from a Normal Population," Biometrika, XVII (1925), Parts 3 and 4, 364-87.

_______, The Methods of Statistics. Second edition, revised; London: Williams and Norgate, Ltd., 1937. 280 pp.

Wilks, S. S., Mathematical Statistics. Princeton University Press, 1944. 284 pp.

Yule, G. Undy and M. G. Kendall, An Introduction to the Theory of Statistics. Eleventh edition, revised; London: Charles Griffin and Company, Ltd., 1937. 570 pp.